

**EXPLORING PHYSIOLOGICAL MEASURES FOR PREDICTION AND
IDENTIFICATION OF THE REDLINE OF COGNITIVE WORKLOAD**

A Thesis

by

CAROLINA RODRIGUEZ PARAS

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Thomas K. Ferris
Committee Members,	Erick Moreno Centeno
	Larry Gresham
Head of Department,	Cesar O. Malave

August 2015

Major Subject: Industrial Engineering

Copyright 2015 Carolina Rodriguez Paras

ABSTRACT

Research suggests that physiological measures such as breath rate (BR), heart rate (HR), heart rate variability (HRV), skin conductance response (SCR), and electroencephalography (EEG) tend to be real-time indicators of mental workload, which are related with increases in the sympathetic nervous system. With increased cognitive workload, these physiological measures tend to change, until a plateau is reached. At this point, performance will decrease, as the workload imposed on the user exceeds their mental capacity to perform the task. This occurs when the user reaches their cognitive redline of workload. Performance will start to decline or decline more steeply at this point, as task demand imposed by the tasks is greater than the mental capacity.

This thesis seeks to understand the underlying patterns reflected in the physiological data that can potentially be used as real-time indicators of the cognitive redline of workload. The study involved use of the Multi-Attribute Task Battery II (MATB-II) to manipulate workload. Subjective measures and performance were taken at the end of every scenario, while physiological measures (BR, HR, HRV, SCR, and EEG), and performance were analyzed to determine the cognitive redline. Results found subjective measures to be responsive to workload change, while heart rate variability seems to be the best physiological measure to respond to mental workload. EEG and SCR proved to also be reliable predictors.

DEDICATION

To my parents, and brother, for all their help and encouragement.

ACKNOWLEDGEMENTS

This project would not have been completed by myself, so I would like to thank all these people for their invaluable help along the course of the project.

I would like to thank my committee chair, Dr. Ferris, for giving me the opportunity to start research. I would like to extend my sincere gratitude to Dr. Ferris and my committee members, Dr. Moreno Centeno and Dr. Gresham, for their continuous support and encouragement throughout the course of this project.

I would like to acknowledge the help of the students from the Human Factors and Cognitive Systems Laboratory at Texas A&M for all their help and support. In particular, Shiyang Yang and Katie Tippey for guiding me, especially during the beginning of the project.

I would also like to thank my parents, Patricio and Diana, and my brother, Patricio, for encouraging me to pursue a Master's Degree at Texas A&M University, and for always being there for me, and their support during the course of this project.

NOMENCLATURE

BR	Breath Rate
EEG	Electroencephalography
HR	Heart Rate
HRV	Heart Rate Variability
GSR	Galvanic Skin Response
MATB-II	Multi-Attribute Task Battery II
NASA-TLX	NASA – Task Load Index
SC	Skin Conductance
SCR	Skin Conductance Response
SCL	Skin Conductance Level
SSSQ	Short Stress State Questionnaire
WPI	Workload Profile Index

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
NOMENCLATURE.....	v
LIST OF FIGURES.....	viii
LIST OF TABLES	ix
LIST OF GRAPHS.....	x
INTRODUCTION.....	1
Literature Review	3
Objective.....	9
Expectations and Hypotheses	11
Contributions of this Research.....	12
METHODOLOGY	13
Apparatus and Devices	13
Physiological Measures Devices	13
Software to Manipulate Mental Workload	18
Questionnaires	23
Participants.....	27
Study Design.....	27
Procedure	28
RESULTS AND EVALUATIONS.....	30
Physiological Measures	30
Subjective Measures	34
Performance.....	35
Relations Between Variables.....	38
Analysis	38

CONTRIBUTIONS AND IMPACT	40
FUTURE RESEARCH	41
CONCLUSION	42
REFERENCES	43

LIST OF FIGURES

	Page
Figure 1: Two areas of task performance (adapted from Wickens et al., 2012)	5
Figure 2: BioHarness3 chest strap.....	15
Figure 3: NeuroSky MindWave	17
Figure 4: ShimmerGSR system.....	18
Figure 5: MATB-II screen, adapted from (Santiago-Espada et al., 2011)	19
Figure 6: NASA-TLX, adapted from (Santiago-Espada et al., 2011).....	24
Figure 7: WPI used during the study.....	25

LIST OF TABLES

	Page
Table 1: Methods Used to Measure Workload.....	7
Table 2: Physiological Measures and their Relationship to Workload	13
Table 3: MATB-II Tasks	20
Table 4: MATB-II Level Description	22
Table 5: Counterbalance Experiment	28

LIST OF GRAPHS

	Page
Graph 1: Breath Rate	30
Graph 2: Heart Rate.....	31
Graph 3: Heart Rate Variability	32
Graph 4: Skin Conductance Response	33
Graph 5: EEG - Alpha Wave.....	33
Graph 6: Workload Profile Index.....	34

INTRODUCTION

It is important to define acceptable levels of workload for the human operators, especially when these are tied to safety and performance in the workplace. Workload is present not only in the industry, such as in healthcare, aviation, and industrial production lines, but in day-to-day daily activities, such as driving (Kohlmorgen et al., 2007). Under high workload, performance tends to decrease. It is possible to prevent performance loss when the user is under high workload and stress by introducing additional resources (Robert & Hockey, 1997). These additional resources can come as additional workers, or automation, to help ease the workload of the human operator. Being able to reliably assess human workload in real-time setting can help identify the point before the user's capacity is reached and exceeded, keeping performance at an optimal level, and maintaining employee safety in the workplace.

A clear example of high workload and stress affecting performance is in the healthcare industry. Physicians who work under these conditions can also see increased risks for occupational health hazards (Homborgh & Engels, 2005). Similarly, nurses who work under high workload have subjectively reported providing lower quality of care, and they experience higher fatigue and stress. An advantage of being able to modulate workload is improving the quality of working life (Gurses, Carayon, & Wall, 2009).

Another area that would greatly benefit from being able to assess workload in real-time setting is the aerospace industry. Automation introduction in aircraft has created a greater need to monitor workload in real-time. Pilots need to be aware of all the

automation systems in the aircraft, as they need to be able to take command in case of a failure of one of the systems (Battiste & Bortolussi, 1988). Having high workload can affect total system performance, and there are costs associated with this, such as degradation in human performance on concurrent tasks and automation false alarms or misses (Dixon & Wickens, 2006).

High workload not only affects performance, but it can endanger the users who are under high workload and the persons they interact with as well. This creates a need to understand rises in cognitive workload, and detecting the maximum capacity of the user in real-time in order to provide additional help. Research is trying to expand the mental workload analysis into developing new human-machine interfaces, and increasing satisfaction, efficiency and safety in the workplace (Rubio, Díaz, Martín, & Puente, 2004). Three main groups of methods commonly used to measure mental workload are employed for this purpose: 1) subjective ratings of workload, 2) performance measures used to represent workload, and 3) physiological measures that have been shown to correlate with mental workload (Brookhuis & de Waard, 2010; Di Stasi, Antolí, Gea, & Cañas, 2011; Vidulich & Tsang, 2012; Wickens, Hollands, Parasuraman, & Banbury, 2012; Wilson, Caldwell, & Russell, 2007).

Physiological measures, such as heart rate and breath rate, are already being monitored in real-time to determine stress levels that affect performance, as in the Indy 500 car race (AP, 2015). Having the workload levels at the optimal level can insure adequate performance, as the operator will not suffer from under- or overload (Kohlmorgen et al., 2007). Having a human operator underloaded or overloaded can

contribute to performance decrements (Jeong & Biocca, 2012), as explained by the Yerkes-Dodson law (Yerkes & Dodson, 1908).

A possible solution would be to monitor in real-time the amount of cognitive workload that the user is experiencing at any given moment. Preferably, to monitor the threshold where performance decrement starts, and the mental capacity is reached and exceeded, which is otherwise known as the cognitive redline of workload (Grier et al., 2008; Wickens, 2008). This thesis will seek to investigate the possibility of using physiological measures as real-time indicators of mental workload, and determine their validity as indicators of the cognitive redline. Physiological measures were chosen as they are the only form of workload assessment that can provide results in real-time without task disruption, as well as being used to determine safety risk, and where possible, drive adaptive systems. These methods are less obtrusive to the worker than subjective methods, where the worker would need to stop the task in order to fill a questionnaire. In addition, an analysis will compare said physiological methods to define which are more reliable as indicators of the redline of cognitive workload.

Literature Review

No single definition exists to describe mental workload, and authors tend to agree that it is very difficult to define, depending on the context in which the term is used, such as performance or task demand. Performance of the human operator declines in a consistent manner when the mental capacity is exceeded by the mental workload (Wang, Hope, Wang, Ji, & Gray, 2012). This point is also referred to as the cognitive “redline”

of workload (Wickens, 2008), which has turned into a high-interest point for the human factors area in recent years. In terms of task demand, mental workload is described as the amount of mental resources that are needed to complete either a single task or multiple tasks (Wickens, 2008), and by the task demands that are imposed on the brain's limited capacity to process information (Wickens et al., 2012).

Two types of tasks that impose on the brain's limited resources are classified according to their demand level, as seen in Figure 1: Two areas of task performance (adapted from Wickens et al., 2012). The first one corresponds to having less task demand than the resources that are available, which results in residual capacity. The second occurs when the demand surpasses the brain's limited ability to process information, which potentially leads to performance decrements. The cognitive redline of workload is defined as the difference that occurs between these two areas of task demand (Wickens, 2008). In other words, the redline is the threshold where performance loss either begins, or if already present, continues to decrement in an escalated way as task demands surpass the brain's available resources (Grier et al., 2008). As explained by the graph, the cognitive redline would occur at the intersection where the primary task performance and the resources supplied meet. At this point, the reserve capacity comes to an end, and task demand exceeds the user's available resources.

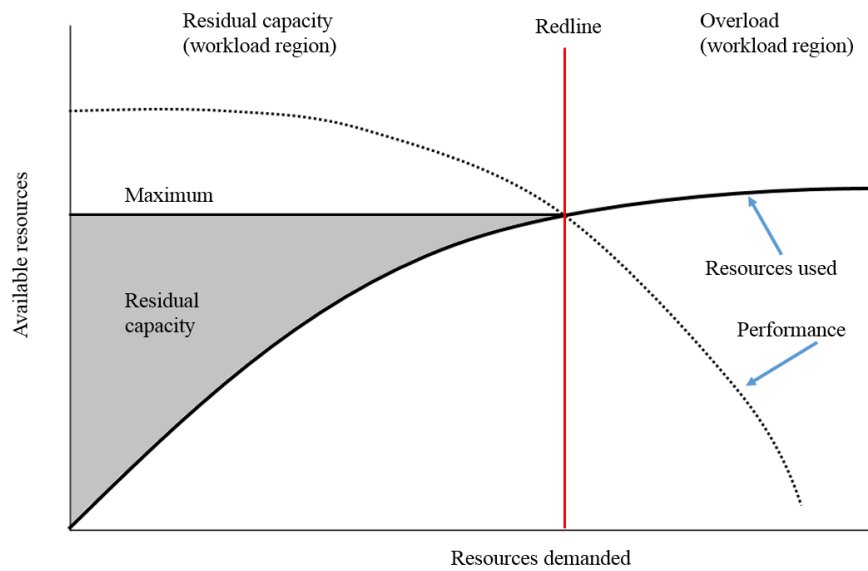


Figure 1: Two areas of task performance (adapted from Wickens et al., 2012)

Mental workload has been previously studied, suggesting that it is possible to reach the cognitive redline, as demonstrated by the correlation between increasing mental workload and changing physiological measures (increasing or decreasing, depending on which physiological measure). Increases in mental workload cause some physiological measures, such as breath rate, skin conductance and heart rate, to reach a point where they no longer change: a plateau (Mehler, Reimer, Coughlin, & Dusek, 2009). These physiological measures are dependent upon the nervous system response. The autonomic nervous system is responsible for activities and reactions that the person is not aware of, such as having cardiac responses to a stressor, and maintaining homeostasis (Waterhouse & Campbell, 2014). It is divided into two components: the sympathetic nervous system and the parasympathetic nervous system (Waterhouse & Campbell, 2014). The sympathetic nervous system is responsible for the fight-or-flight

response. Once the sympathetic nervous system is activated by stress, it stimulates all the functions that are controlled by it, such as cardiac output, sweating, blood pressure and sugar (Waterhouse & Campbell, 2014). Workload increases the activity in the sympathetic nervous system, while the parasympathetic nervous system tends to decline with added workload (Cinaz, Arnrich, La Marca & Troster, 2013). Similarly, event-related potential (ERP) shows asymptotes when the capacity of the visual working memory is exceeded (Vogel & Machizawa, 2004). When this occurs, and the mental working capacity is exceeded, the learning ability degrades (Coyne, Baldwin, Cole, Sibley, & Roberts, 2009; Paas, Renkl, & Sweller, 2004). Degradation of multitasking performance and demonstrate that physiological measures might not be the only possible indicators of the cognitive redline, as multitasking performance tends to decrease when the cognitive redline is reached (Grier et al., 2008). The plateau related to BR, HR, HRV, GSR, and the ERP asymptotes could be the possible physiological measures that reflect the cognitive redline of workload, which will be analyzed with further detail in this thesis.

Determining the cognitive redline using physiological measures in real-time can help to reliably assess the human operator's level of workload in environments that have a high workload demand. Knowing the level of workload can certainly prove to be beneficial, especially when the user is at or near their individual cognitive redline. Mental workload tends to be assessed during the first few stages of system design, such as system design and evaluation, especially in settings that are known to induce high workload (Vidulich & Tsang, 2012). Current research is aiming at developing novel and

effective displays and tools that can be used as aids in high multitasking environments with the sole intention of bringing the user's workload levels to acceptable levels (Grier et al., 2008). Safety and performance can be increased in high-workload environments if the mental workload is managed effectively (Krehl & Balfe, 2014). High-workload environments not only occur in the industry setting, as it can be encountered in daily routing activities, such as driving. For example, many traffic accidents occur when the driver has improper workload levels. Low workload levels, also known as vigilance problems, and high workload levels or stress can be very problematic (Brookhuis & de Waard, 2010), especially with new and inexperienced drivers.

Being able to accurately assess workload might make it possible to quantitatively define and predict the point at which the human operator reaches their cognitive redline. Each of these groups of methods poses limitations as well as benefits particular to each of the methods, making each one suitable for workload assessment under different circumstances. A summary of the three different groups of methods (including examples used and their citations), is presented in Table 1: Methods Used to Measure Workload.

Table 1: Methods Used to Measure Workload

Method	Example	Citation
Physiological methods	Breath rate (BR)	(Roscoe, 1992)
	Heart rate (HR)	(Solovey, Zec, Garcia Perez, Reimer, & Mehler, 2014; Veltman & Gaillard, 1996)
	Heart rate variability (HRV)	(Karim, Hasan, & Syed, 2011)
	Skin conductance response (SCR)	(Mehler et al., 2009)

Table 1: Continued

Method	Example	Citation
Physiological methods	Electroencephalography (EEG)	(Berka et al., 2007)
Subjective methods	NASA-Task Load Index (NASA-TLX)	(Hart & Staveland, 1988)
	Dundee Stress State Questionnaire (DSSQ)	(Helton, 2004; Matthews et al., 2006)
	Workload Profile Index (WPI)	(Rubio et al., 2004)
Performance	Performance decrements	(Grier et al., 2008; Wang et al., 2012)

Subjective methods are used to measure mental workload, which are non-intrusive, and easy to implement. However, these methods take into consideration the assumption that the human users can perceive the effort that is related to more power expense during a certain task (Rubio et al., 2004).

Another indicator of mental workload imposed by the task can be measured through task performance. Performance decrements occur due to increases in mental workload that exceed the brain's limited mental capacity (Wang et al., 2012). Dissociation is possible when measuring performance, which means that a performance decrement is not always a good indicator of cognitive workload. Furthermore, there is challenge to differentiate effort from workload (Hancock, Williams, Manning, & Miyake, 1995). The possible solution to solve this problem is to insert a secondary task to act as a probe that could potentially interfere with the primary task performance and workload (Hancock et al., 1995). The other possible problem with performance is that it could also reflect other states of the user, such as boredom. This can happen when the user is underloaded and thus is not paying attention to the task at hand, as explained by

the Yerkes-Dodson law. This law explains that the pattern for cognitive performance takes the shape of an inverted U-shaped pattern, where the maximum performance is obtained through optimal levels of workload, and is affected by overloading or underloading of the user (Jeong & Biocca, 2012; Yerkes & Dodson, 1908).

The last group of methods to analyze changes in mental workload are physiological measures, which have been studied extensively, often used in combination with subjective and/or performance measures. Using physiological measures as indicators of mental workload provides several advantages over the other two methods, such as the ability to continuously measure workload levels in real time, with little to no intrusion in the task (Prinzel, Freeman, Scerbo, Mikulka, & Pope, 2000). The major drawback is that data can be very noisy, depending on the quality of the devices used to collect the data, as well as having difficulty to isolate the effects of cognitive workload from other affecting factors, such as fatigue or stress, that could affect physiology.

All three groups of methods will be studied in this thesis, as explained in the next section. The purpose is to be able to use subjective methods and performance to validate physiological measures as true indicators of the cognitive redline of workload.

Objective

Few studies have pursued research in regards to the cognitive redline of workload, leading to few existing definitions of the cognitive redline. One reason is that mental workload is difficult to quantify through physiological measures, making it more difficult to find the cognitive redline. This project will use a combination of the three

mental workload assessment methods (physiological measures, performance, and subjective measures) to determine if physiological measures can be used as indicators of the cognitive red line of workload. If this is possible, further analysis will be performed to describe which physiological measures are more reliable in predicting the redline of cognitive workload.

It can be very beneficial if we can find the ability to detect, in real-time or retrospectively, when an operator has approached or exceeded their cognitive redline. This benefit can be applied during the design of task environments that impose high task demands. Some of these environments include the aerospace industry (aircraft cockpits), hospitals (surgery rooms, emergency rooms, and anesthesiology), warehouses (logistics and transportation), nuclear reactors, and the driving environment.

Another benefit to detecting the cognitive redline is that it can aid in research to develop displays and tools for jobs that require multitasking. Knowing the redline of the users can help establish acceptable workload levels for the users (Grier et al., 2008), which can result in increased workplace safety for the user. Task safety and performance in these high-workload environments depends mainly on effectively managing workload (Krehl & Balfe, 2014). Having a clear understanding of the cognitive redline of workload can lead to knowing what levels of workload can be deemed safe to prevent accidents due to high workload.

Identifying when a human operator might be approaching their cognitive redline can allow the system to effectively monitor the performance in the multitasking setting. At this point, assistance can be provided to the human to reduce performance

decrements. This assistance can be in the form of another worker, or even introducing automation with an instrument approach to aid the worker and not remove him completely from the workplace.

This thesis will provide both theoretical and applied contributions. The applications of the knowledge gained will lead to determining the most reliable real-time indicators for the redline of cognitive workload. Designers for data-rich environments can incorporate this information to potentially predict when operators may become overloaded and thus when their performance will suffer and their safety might be compromised. The theory is to contribute to the knowledge base with regard to the relationship between physiological indicators, cognitive capacity, and performance.

Expectations and Hypotheses

Two hypotheses will be tested in this thesis. The redline of cognitive workload can be determined by physiological measures (such as BR, HRV, HR, SC, and EEG) by identifying the point in which they stop increasing or decreasing (increasing for BR, HR, SC, and decreasing for HRV and EEG alpha wave). These changes in physiological measures will be detected with the statistical analysis, where the variance will be taken between the 5 difficulty levels. The variance will be greater for the first few levels, but the variance should be minimal between the most difficult levels. This leads to the development of the hypotheses. **Hypothesis 1:** Physiological measures changes (decreasing or decreasing, depending on the measure) will correlate with level increase for the first few levels, while for the last levels marginal correlation with increased level

will be found (when the redline is reached). **Hypothesis 2:** Performance decrements will correlate significantly with difficulty levels.

Contributions of this Research

It is difficult to find literature information about the cognitive redline, as it rarely gets mentioned in literature. Many studies have looked at performance decrements with increased workload, and analyzed changes in physiological measures due to increased workload, but few include discussion that offers insight into the concept of the redline (e.g., “plateau” patterns (Mehler et al., 2009) or asymptotes of physiological measures (Vogel & Machizawa, 2004)). This thesis seeks to go further, and to actually demonstrate the existence of said redline of cognitive workload through physiological measures plateaus and performance decrements. The second objective of this thesis is to determine how reliable physiological measures are as predictors of workload, and which of these measures the best indicators are.

METHODOLOGY

Apparatus and Devices

Several apparatus and devices were used throughout the study, ranging from physiological devices to questionnaires. Physiological variables were measured using three devices (BioHarness3, ShimmerGSR and NeuroSky MindWave), which detect and measure the physiological variables in real-time. Subjective measures of workload were obtained through questionnaires, such as WPI and NASA-TLX, while mental workload levels were manipulated using a software (Multi-Attribute Task Battery-II).

Physiological Measures Devices

Several physiological measures will be analyzed in this project, as summarized in Table 2: Physiological Measures and their Relationship to Workload, due to the fact that it is not very clear which physiological measures are the best indicators of the cognitive redline. All physiological measures are recorded so that more types of data can be analyzed and then compared to subjective methods and performance.

Table 2: Physiological Measures and their Relationship to Workload

Physiological Measure	Relationship with increased workload	Device used to measure in this study
Breathing rate (BR) or respiratory rate (RR)	Increases	BioHarness3
Heart rate (HR)	Increases	BioHarness3

Table 2: Continued

Physiological Measure	Relationship with increased workload	Device used to measure in this study
Heart rate variability (HRV)	Decreases	Calculated from R-to-R interval provided by BioHarness
Skin conductance (SC) or galvanic skin response (GSR)	Increases	ShimmerGSR
Electroencephalography (EEG), alpha wave (8-13 Hz)	Decreases or increases	MindWave

A total of three wearable devices were used in the study to record physiological measures. BioHarness3 (see Figure 2: BioHarness3 chest strap) measures BR and HR, and also provided the R-R interval needed to compute HRV. It consists of a biomodule with microcontroller and sensor, which is attached through an adjustable chest strap to the left side of the torso, near the heart. The chest strap has to be wet before each use to increase the contact between skin and sensors. This device has been validated to use in experimental studies (ZephyrCorporation, 2008).

Respiratory rate (RR), or breath rate (BR) are among the several respiratory variables that have been studied in previous studies described in literature to study mental workload. Mean BR is 12 breaths per minute for a healthy adult, and tends to increase with added mental workload (Roscoe, 1992).

Heart rate (HR) has been demonstrated by studies as an effective indicator of mental workload (Solovey et al., 2014). HR is controlled by the autonomous nervous

system, which is broken down into the sympathetic and parasympathetic components. The sympathetic nervous system increases HR, while the parasympathetic decreases it. The sympathetic nervous system responds to stressors by affecting cardiac output and sweat production (Waterhouse & Campbell, 2014). Increases in mental workload are reflected by increases in HR (Veltman & Gaillard, 1996).

Heart rate variability (HRV) can be calculated from the peak-to-peak interval (R-R interval), usually in the form of short time intervals between 2 to 5 seconds. HRV occurs to variations that happen between consecutive heartbeats, and several methods can be used to calculate it, such as the frequency-domain or time-domain (Karim et al., 2011). With increased mental workload, HRV will decrease (Veltman & Gaillard, 1996), and has been found to reliably shown the differences in task demands (Cinaz et al., 2013; Galy, Cariou, & Mélan, 2012).



Figure 2: BioHarness3 chest strap

NeuroSky Mindwave served to record the EEG measurements from the participants, as seen in Figure 3: NeuroSky MindWave. This device consists of an adjustable headband, with a sensor located on the left-hand side of the forehead (the FP1 position), which allows it to record EEG frequency bands as well as blinks ("Why locate the sensor at FP1? ," 2014). The device connects through Bluetooth to the computer, providing the data in real-time. The software allows the user to see the data or generate graphs in real-time. A reference electrode is placed on the left ear. The advantages of MindWave over other commercial EEG devices is that it is very easy for the user to wear, since is it an adjustable headband, and does not require connective gel.

ElectroEncepheloGraphy (EEG) records the voltage differences that occur between a reference electrode (which is usually placed on the ear), and the active electrodes that are placed on the head. Several frequency bands are identified in EEG analysis: 0.5-3 Hz (delta), 4-8 Hz (theta), 8-13 Hz (alpha), 13-30 (Beta) and 40 to 50 Hz (gamma) (Mehta & Parasuraman, 2013). The alpha frequency band is lowered by increases in cognitive workload, as the 8-13 Hz frequency represents attenuated activity (Berka et al., 2007; Galy et al., 2012), and the beta and theta frequency bands increase their frequency wave (Stephen H. Fairclough, Venables, & Tattersall, 2005). Alpha wave tends to increase when the user experiences pleasant feelings and has a narrow perceptual awareness (Brown, 1970).



Figure 3: NeuroSky MindWave

ShimmerGSR (see Figure 4: ShimmerGSR system) is a module used to measure SC, from which SCR and SCL can later be derived. Shimmer has 2 electrodes, which connect to an amplifier that gets attached to the participant's wrist through a strap, and transmits the information through Bluetooth. The software used to record the information (ShimmerSense) in real-time also allows for the display of graphs in current-time. The 2 electrodes, located at the end of the two lead wires fasten to disposable isotonic gel electrodes, which attach to the hand. It is common to use the fingers, but to reduce data noise, the electrodes were attached to the palm of the dominant hand of the user in this study. Electrodermal activity (EDA), sometimes referred to as galvanic skin response (GSR), such as skin conductance level (SCL), or skin conductance response (SCR), measures electrical resistance and conductance of the skin. The general skin potential is broken down into 2 components: tonic (level), represented by slow changes, and phasic (response), corresponding to rapid peaks (Boucsein, 2012). The sympathetic nervous system controls the eccrine sweat glands,

which changes the electrical resistance and conductance of the skin (Solovey et al., 2014). Studies have demonstrated that EDA shows asymptotes with increased mental workload (Vogel & Machizawa, 2004), making it one of the possible predictors of the cognitive redline of workload.

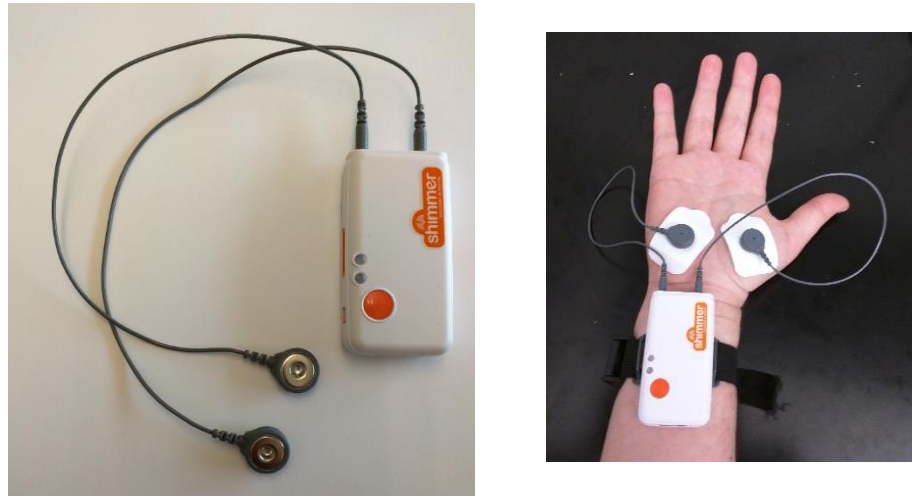


Figure 4: ShimmerGSR system

Software to Manipulate Mental Workload

The aforementioned devices are used to evaluate mental workload, which can be influenced in order to evaluate performance. A computer software that offers this capability is the Multi-Attribute Task Battery II (MATB-II, as seen in Figure 5: MATB-II screen), developed by NASA (Santiago-Espada, Langley Research, United States. National, & Space, 2011). Previous studies have made use of this software to induce different levels of mental workload (Chiappe, Conger, Liao, Caldwell, & Vu, 2013; Wang et al., 2012). MATB-II contains four different tasks: system monitoring, resource

management, communication and tracking (Santiago-Espada et al., 2011). Workload is manipulated by changing the number of incidences or events in a given set of tasks.

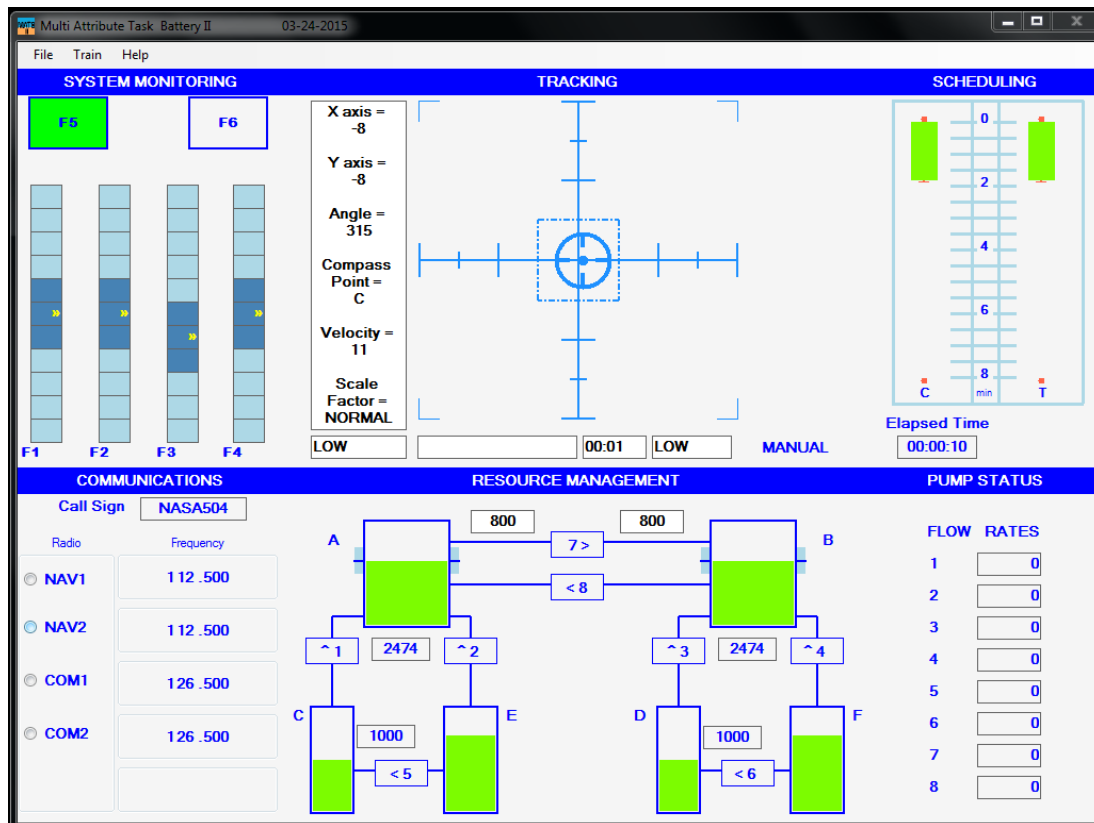


Figure 5: MATB-II screen, adapted from (Santiago-Espada et al., 2011)

The Multi-Attribute Task Battery II (MATB-II) software has been used to manipulate mental workload. The software simulates an aircraft, using tasks similar to those used by aircraft pilots during flight (Comstock & Arnegard, 1992). Physiological measures have been used with MATB-II in several studies with the purpose of studying mental workload (Fairclough et al., 2005; Miyake et al., 2009; Prinzel et al., 2000; Wang

et al., 2012). It has four tasks, which users have to monitor and respond to during the session. These tasks are: communication (COMM), tracking task with the use of a joystick (TRACK), resource management (RESMAN), and system monitoring (SYSMON).

A short description of the four tasks used in the study are described in Table 3:
MATB-II Tasks.

Table 3: MATB-II Tasks

Task	Description
Resource management (RESMAN)	<p>Tanks A and B need to be kept at the optimal level of 2500 +/- 500 by transferring fuel from the other tank reservoirs (C, D, E and F).</p> <p>Tanks E and F are unlimited. Fuel is transferred by pressing the number on each of the pumps through the keyboard or mouse. When the pump is white, it is off. A green pump indicates that it is working, and a red one is malfunctioning.</p> <p>The difficulty is manipulated through how often the pumps fail, as well as how long each failure lasts.</p>
Communications (COMM)	<p>Users listen to several communications, but only need to respond to their call sign: NASA504. They hear the call sign twice, followed by the type of radio and the frequency they have to input.</p> <p>The frequency with which the different communications appear are used to increase workload.</p>

Table 3: Continued

Task	Description
System Monitoring (SYSMON)	<p>This task has two systems: system lights and scales. The lights' normal colors are green and white. The user has to press on them when the green turns off and when the white turns red by pressing on the squares or by pressing the F5 and F6 keys.</p> <p>The scales will be fluctuating in the middle, and the user has to respond when they get near the top or the bottom. The user can respond through the keyboard or with a mouse.</p> <p>Workload can be increased in this task by increasing the number of incidences in which the user has to respond to the tasks.</p>
Tracking (TRACK)	<p>Participants use their dominant hand to control the pointer through the use of a joystick. The pointer has to be kept inside the square. There are 2 ways to increase difficulty: through response and update. Each of these can be set to easy, medium or hard.</p>

The MATB-II levels have varying levels of cognitive workload, based on gradually increasing the iterations and difficulty for each of the tasks (Mehler et al., 2009). The levels were designed through extensive pilot testing in order to have them be equally spaced in terms of difficulty. For the pilot test, a table was created with the number of iterations that each level needed to have in order to have the difficulty be

equally spaced (for example, changing the number of seconds for pump timeouts in resource management, number of communication tasks users had to respond to, number of incidences in system monitoring and changing the update and response in the tracking task). Once the levels were coded, they were pilot tested and modified according to user feedback (which included saying which tasks were easier or more difficult, and using subjective measures, such as NASA-TLX and WPI). They kept being modified until user feedback was consistent with the difficulty level assigned to each. The difficulty depended on the settings specified for every task. For example, the level of the joystick required to complete the tracking task, could be set to easy sensibility during the easy tasks, and increased to high for more difficult levels. Communication task had few calls in the easier levels, with the number of communication events increasing through the progressing levels. Resource management and system monitoring had few cues, which progressively increased. A summary of the difficulty levels is given in Table 4: MATB-II Level Description. The response timeout for resource management is 10 seconds.

Table 4: MATB-II Level Description

Level	Color	TRACK		SYSMON			RESMAN Total Time			COMM	
		Update	Response	Green	Red	Scale	Fail	Fix	(s)	Own	Other
1	Yellow	Low	Low	1	1	1	2	2	10	1	1
2	Purple	Low	Medium	4	4	6	5	5	45	2	2
3	Blue	High	High	6	6	15	5	5	43	5	3
4	Red	High	High	10	10	20	5	5	112	5	3
5	Green	High	High	11	11	22	10	10	108	5	5

Questionnaires

Numerous subjective methods exist to assess mental workload, such as NASA-Task Load Index (NASA-TLX) and Workload Profile Index (WPI). These methods have been constantly used to evaluate operator workload in several systems. The advantages they offer include being non-intrusive, easy to implement, and sensitive to operator workload (Rubio et al., 2004). These methods, even though they can be very good in perceiving the overall workload from the user's perspective, can disrupt the main task unless they are assessed retrospectively, which can then be susceptible to memory effects. This is one of the reasons why physiological measures are often used with subjective methods.

NASA-TLX classifies mental workload into six categories: effort, frustration, performance, mental, physical, and temporal demand (Hart & Staveland, 1988). Once the task is complete, users first provide a rating on a fixed scale (for example, 1 to 100) for each of the six dimensions. To estimate the relative contributions of each dimension to the overall perception of mental workload, a pairwise comparison procedure is then followed, where users select the highest category from each of the 15 pairs. Each of the individual ratings and weightings are informative in themselves, and then a final overall task load index (TLX) is calculated as the sum of each individual rating multiplied by its respective weighting (Fréard, Jamet, Le Bohec, Poulain, & Botherel, 2007; Rubio et al., 2004). Physiological measures have been used in conjunction with NASA-TLX to gauge mental workload (Galy et al., 2012; Kim & Ji, 2013), especially with EEG (Hancock & Szalma, 2003).

NASA-TLX (NASA-TLX) appears at the end of every MATB-II level as an electronic survey, as seen in Figure 6: NASA-TLX. Participants complete the NASA-TLX ratings after each scenario of MATB-II, while the last section of NASA-TLX, the pairwise comparison get completed at the end of the study, asking the user to rate them according to the 5 scenarios.

The screenshot shows the NASA-TLX survey interface. It consists of six horizontal rating scales, each with a green slider indicating the current rating. The scales are labeled 'Low' and 'High' (or 'Good' and 'Poor' for Performance). The scales are: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. At the bottom, there are buttons for 'Reset Ratings' and 'Save All'.

Dimension	Rating Scale (Low to High)	Current Rating (Approximate)
Mental Demand	Low to High	7.5
Physical Demand	Low to High	5.5
Temporal Demand	Low to High	8.5
Performance	Good to Poor	5.5
Effort	Low to High	5.5
Frustration	Low to High	2.5

Figure 6: NASA-TLX, adapted from (Santiago-Espada et al., 2011)

The Workload Profile Index (WPI) is also completed at the end of every MATB-II scenario. WPI takes advantage of Wickens' Multiple Resource Theory model, as users rate the tasks according to how each task loads individual resources, which are defined by the dimensions of processing stage (Perception/Cognition or Response), sensory modality (Vision or Audition or Haptic), and processing code (Verbal/Symbolic or Spatial/Analog) (Wickens, 2008). Some concerns have arisen in regards to explaining

the variations that are seen in time-sharing efficiency, even though this method is successful in determining interference that occurs between tasks in a multitasking setting (Phillips & Boles, 2004). In WPI, users get a list of tasks completed from a multitasking setting, and they compare these tasks against the Multiple Resource Theory Model. The ratings go on a scale of 0 to 1 (0 for no usage and 1 for full usage) (Rubio et al., 2004). For the study explained in this thesis, the rating scale was changed from 0 to 5 (0 for no engagement and 5 for full engagement). The parts of the Multiple Resource Theory Model that were not used during the study were also removed to make it easier for participants to respond to the questionnaire. A sample questionnaire is seen in Figure 7: WPI.

Please fill an integer number from 0 to 5 into the box.

0 = No engagement at all 5 = full engagement.

Yellow	Workload Dimensions						
	Stage of Processing		Code of Processing		Input		Output
	Perceptual/ Cognition	Response	Spatial	Verbal	Visual	Auditory	Manual
MATB-II Task							

Figure 7: WPI used during the study

The Short Stress State Questionnaire, similarly to the Dundee Stress State Questionnaire (DSSQ), measures stress through three factors: 1) task engagement, 2)

distress, and 3) worry (Helton, 2004; Matthews & Campbell, 1998; Matthews et al., 2006). Task engagement refers to the motivation, energy and amount of concentration needed to accomplish the task. Distress are all the negative effects that arise due to the user knowing they lack control of the task, and worry encompasses all the negative thinking styles (Matthews et al., 2006). The test is taken in two parts, before and after completing the task. The Dundee Stress State Questionnaire has been validated as effective to assess stress levels, but having 90 questions in each portion of the test (pre- and post-task), the results can be affected by participants' fatigue (Helton, 2004). The main advantage of SSSQ is that it only uses 24 questions for each portion of the test, providing results that are very similar to DSSQ (Helton, 2004). SSSQ has been used in conjunction with NASA-TLX to gauge changes in stress due to the experimental setup along with NASA-TLX (Matthews & Campbell, 2010).

For the baseline and resting periods, participants used noise-cancelling earphones to listen to nature sounds while doing breathing exercises. The Paced Breathing App, which was developed by TrexLLC, was used for the breathing exercises (TrexLLC, 2014). This app was used for this purpose, as studies have shown that doing breathing exercises with a ratio of 1 inhalation to 2 exhalations reduces physiological measures, such as heart rate and blood pressure (Modesti et al., 2010). This decreases the average breath per minute from 12, which is the normal average for a healthy adult (Roscoe, 1992), to 6 breaths per minute. The app shows a graph that indicates the user when the inhale, how long to hold the inhalation, when to exhale, and how long to hold the

exhalation. The graph gives the user something to follow and relax, while maintaining their eyes open and not disrupting the EEG recordings.

Participants

The IRB-approved study recruited participants through mass email who are 18 years or older, and who have normal or corrected-to-normal vision. Having no impairments on their hands and arms is a condition to participate in the study.

Study Design

A counterbalance design was chosen based on the five levels of difficulty presented in MATB-II. A counterbalance design will help to reduce the carryover effect and learning curve, as well as any fatigue that participants might feel during the duration of the study. The first 5 levels are balanced, followed by the mirrored table for a total of 10 experimental scenarios, which are repeated 3 times for a total of 30 participants.

The dependent measures include: performance, physiological and subjective measures. The independent variables will consist of the difficulty levels.

Since participants were gazing at the screen when each MATB-II level was opened, each scenario was coded with a non-descriptive color to make participants unaware of the imposed workload, and thus preventing affecting the subjective ratings. The counterbalanced experimental design is seen in Table 5: Counterbalance Experiment, with the second table having the color equivalent of every level.

Table 5: Counterbalance Experiment

1	2	5	3	4	Y	P	G	B	R
2	3	1	4	5	P	B	Y	R	G
3	4	2	5	1	B	R	P	G	Y
4	5	3	1	2	R	G	B	Y	P
5	1	4	2	3	G	Y	R	P	B
4	3	5	2	1	R	B	G	P	Y
5	4	1	3	2	G	R	Y	B	P
1	5	2	4	3	Y	G	P	R	B
2	1	3	5	4	P	Y	B	G	R
3	2	4	1	5	B	P	R	Y	G

Procedure

The IRB-approved study consisted of participants arriving to the lab, and then they were given a short explanation of the study and the devices used. After the explanation, they were given the consent form, and a background questionnaire to determine if they have any previous experience with flight simulators, or if they had consumed caffeine before the study, which affects physiology. The first portion of the SSSQ is also completed during this stage.

Participants were then fitted with the devices to gather physiological measures data: BioHarness3, ShimmerGSR, and MindWave. Once the devices were connected through Bluetooth to the computer, the baseline was recorded. Participants had to use the Paced Breathing App while listening to nature sounds through noise-cancelling earphones for 10 minutes. They were asked to relax, and to keep their eyes open by following the pattern on the app.

After the 10-minute baseline, participants completed a training, which most effectively reduces the likelihood of observing significant learning effects during the

study (Prinzel et al., 2000). There were 2 training scenarios, with difficulty similar to the medium and hard MATB-II scenarios. Each of these training scenarios had a duration of 3 minutes. They were explained how the software works before proceeding complete the training. They also received helpful comments, and information about missing cues while they were performing the scenarios so that the participants could grasp a better understanding of how the software operates.

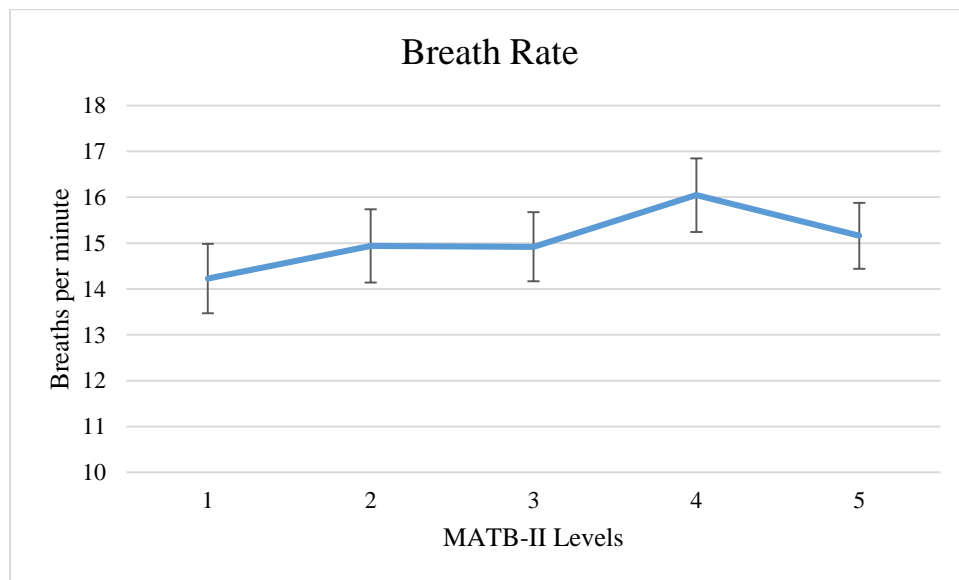
After training, the study proceeded to repeat the following procedure for each of the MATB-II scenarios (5 total). A 3-minute resting period was done, similar to baseline, where participants continued using the Paced Breathing App (TrexLLC, 2014) and listening to nature sounds through noise-cancelling earphones. This was followed by a MATB-II experimental level. After the level was completed, participants completed subjective measures of workload. NASA-TLX is embedded on the MATB-II software, so it appears immediately on the screen, followed by WPI. This same pattern was repeated for the reminding four levels. After the fifth level is complete, participants completed 2 additional questionnaires: the SSSQ post-task and the pairwise comparisons for NASA-TLX.

RESULTS AND EVALUATIONS

A total of 30 participants ($n = 30$) were recruited to perform the study (mean age = 24.26, 12 females and 18 males). All data were analyzed using SAS 9.3, where a within-subjects ANCOVA was performed using the Proc Mixed function in order to reduce maximum likelihood estimates. An $\alpha = 0.05$ significance level was used and standard t-tests were also employed for post-hoc analysis in order to determine the differences between the levels.

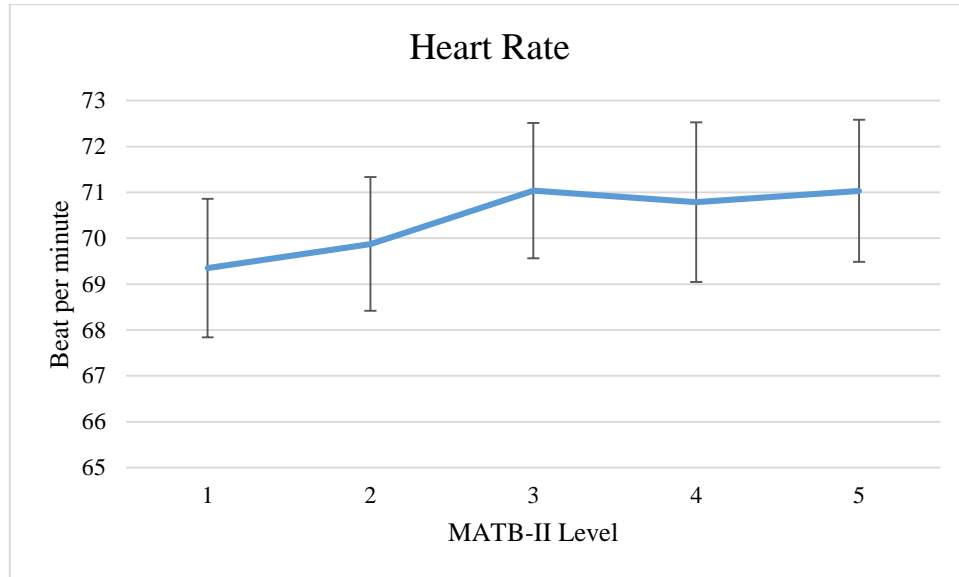
Physiological Measures

Breath rate did not show a significant difference between the workload levels imposed by MATB-II and the breaths per minute ($F(4, 37) = 1.97$, $p = 0.1034$).



Graph 1: Breath Rate

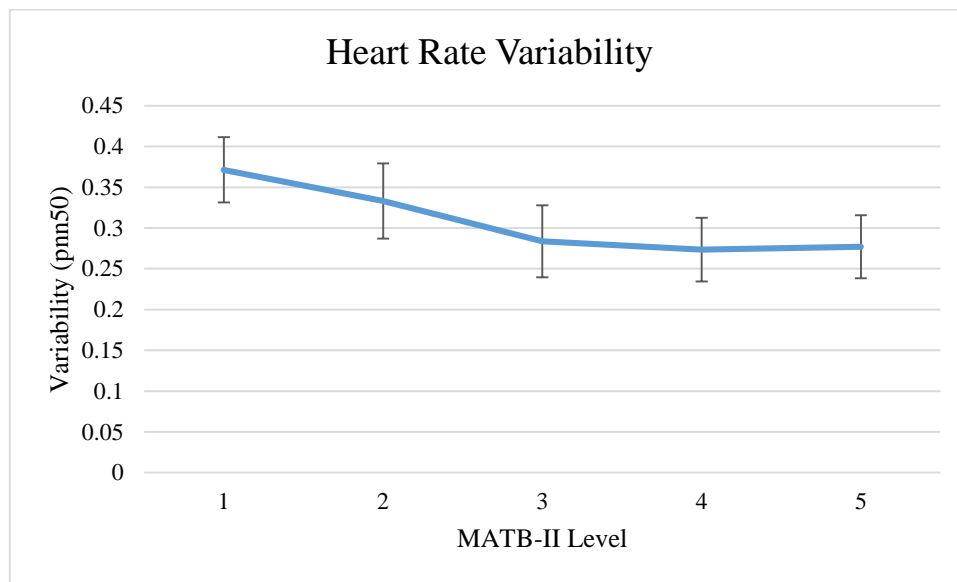
Heart rate does not show a significant difference between the workload levels ($F(4, 38)$, $p = 0.1068$).



Graph 2: Heart Rate

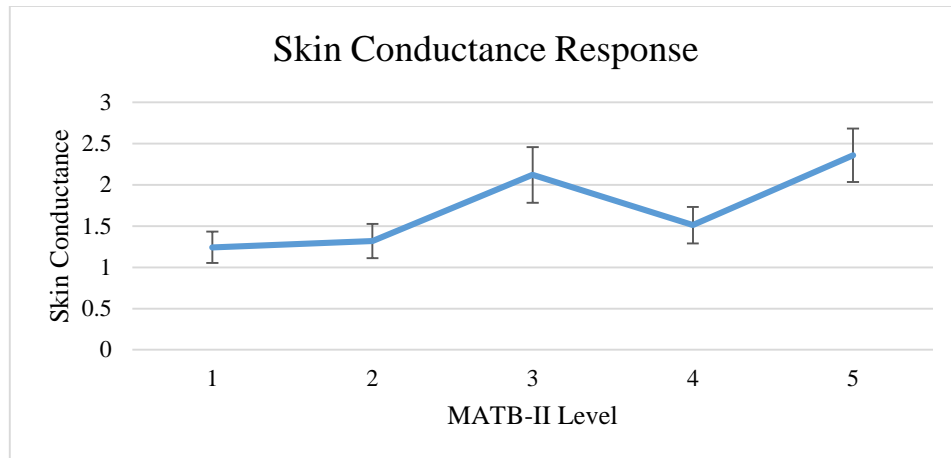
Heart rate variability shows a significant difference for the MATB-II levels ($F(4, 34) = 5.47$, $p = 0.0005$). The pnn50 was used to compute these results, which is the percentage of the heart rate intervals that have a variance greater than 50 ms based on the previous interval (Cinaz et al., 2013). Post-hoc results using Tukey-Kramer show that level 1 is not significantly different from level 2 ($p = 0.6073$), but significantly different from levels 3, 4 and 5 ($p = 0.0084$, $p = 0.0027$, and $p = 0.0043$). Level 2 is not significantly different from levels 3, 4 and 5 ($p = 0.2907$, $p = 0.1486$, and $p = 0.1984$). Level 3 is very similar to level 4 and 5 ($p = 0.9968$, $p = 0.9996$), and level 4 is near-

identical to level 5 ($p = 0.9999$). The significant difference seen in between levels 1 and 2 gives way to levels 3, 4, and 5 being near identical, showing how the HRV no longer changed with added workload. This can be identified as a plateau, indicative of the users having reached their cognitive workload.



Graph 3: Heart Rate Variability

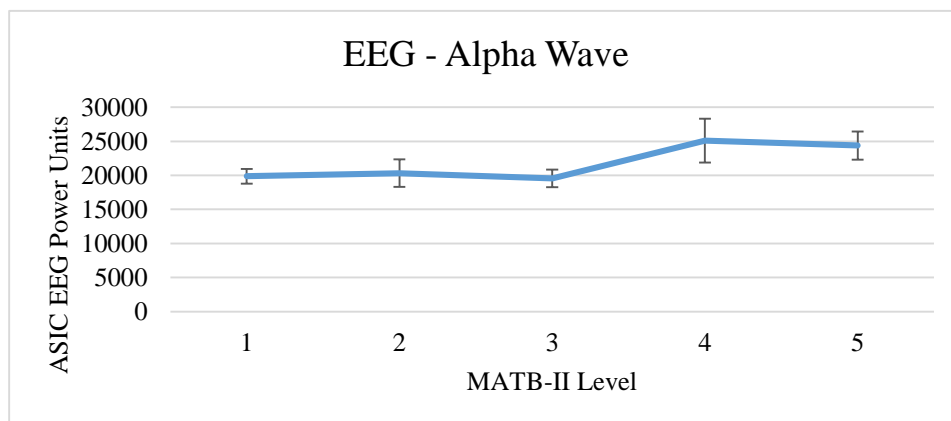
Skin conductance response shows significant results ($F(4, 31) = 3.91$, $p = 0.0057$). Tukey-Kramer post-hoc results show level 1 is not significantly different from all the other levels. Similarly, level 2 shows no significant difference from levels 3 and 4, but it is differently different from level 5 ($p = 0.0151$). All the other levels showed no significant difference among them.



Graph 4: Skin Conductance Response

For EEG, results showed a significant difference ($F(4, 37) = 2.69, p = 0.0348$).

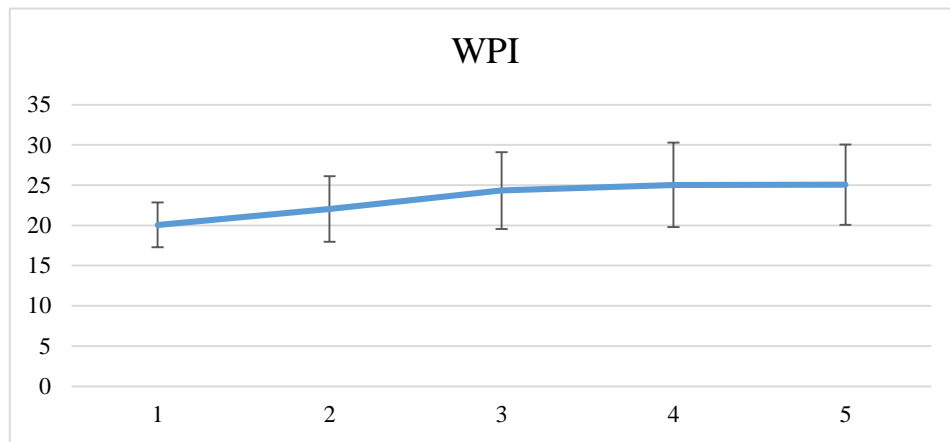
Post-hoc analysis using Tukey found that level 1 is not significantly different from levels 2, 3, but slightly different from levels 4 and 5 ($p = 0.9996, p = 0.9999, p = 0.1605, p = 0.2932$). Level 2 is very similar to level 3 ($p = 0.9973$), and slightly different from levels 4 and 5 ($p = 0.1201$ and $p = 0.2303$). Level 4 is nearly identical to level 5 ($p = 0.9978$).



Graph 5: EEG - Alpha Wave

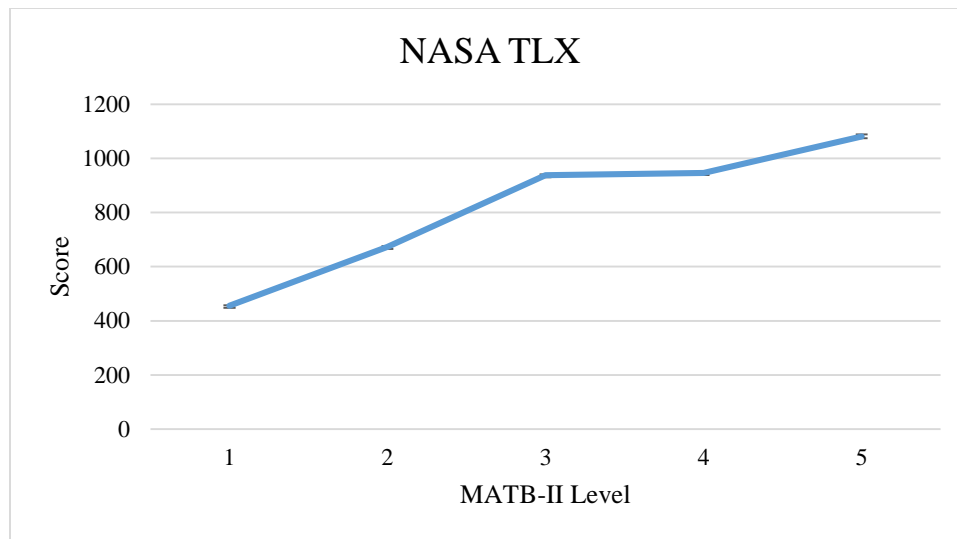
Subjective Measures

WPI was analyzed using Friedman's Test, as it represents rankings in a discrete data set. The test proved WPI to be a significant indicator of mental workload ($\chi^2_r = 35.29$, $df = 4$, $p < 0.001$). Additional post-hoc analysis did not reveal significant difference between the workload levels.



Graph 6: Workload Profile Index

NASA-TLX was analyzed using a within-subject ANOVA across all the five difficulty levels of MATB-II. NASA-TLX was found to be a significant indicator of mental workload ($F(4, 37) = 77.49$, $p < 0.001$). Post-hoc results using Tukey found levels 1 and 2 to be significantly less difficult than all the other levels ($p < 0.001$). Level 3 was not significantly different from level 4 ($p = 0.9995$), but significantly different from 5 ($p = 0.0049$). Level 4 was significantly different from level 5 ($p = 0.0097$).



Graph 7: NASA-TLX

The SSSQ change score was calculated according to the following formula: (Pre-task score – Post-task score)/(Standard deviation of the Pre-task score) (Helton, 2004).

The average stress change score resulted in 0.031, which was found to be not significantly different from 0 based on a Student t-test.

Performance

To calculate the user performance in MATB-II, many studies have looked at performance for every individual task. In particular, studies have looked at the root-mean square deviation from the center target for the tracking task, mean deviation from the required value of units required in both tanks (2500 units), and on the monitoring task performance was dependent on the mean reaction time, number of misses and number of false alarms (Fairclough & Venables, 2006; Stephen H. Fairclough et al., 2005). A more comprehensive formula was used to combine the four different tasks into

a single performance metric in order to find the correlation between HR, performance and task load (Splawn, 2013). Since the software used (MATB_AF) slightly differs from MATB-II, the formula provided could not be applied directly to the performance metrics.

The formula used for system monitoring task is the same one used by Splawn (2013).

$$Xi = \frac{(R_t C) + N_t T}{ET}$$

Where,

R_t = Response time

C = Number of correct responses

N_t = Number of timeouts during the scenario

T = Time (in seconds) needed for the timeout to occur

E = Total number of events in the scenario.

For this task, any additional keystrokes that were not involved in a response were ignored.

Splawn (2013) uses formula 1 for system monitoring and communication. In this case, the formula was modified to accommodate all the different possibilities of correct and incorrect responses that are possible for the communication task.

$$X_i = \frac{(R_t C) + (I_t N_t) T}{ET}$$

Where,

R_t = Response time

C = Number of correct responses

N_t = Number of timeouts during the scenario

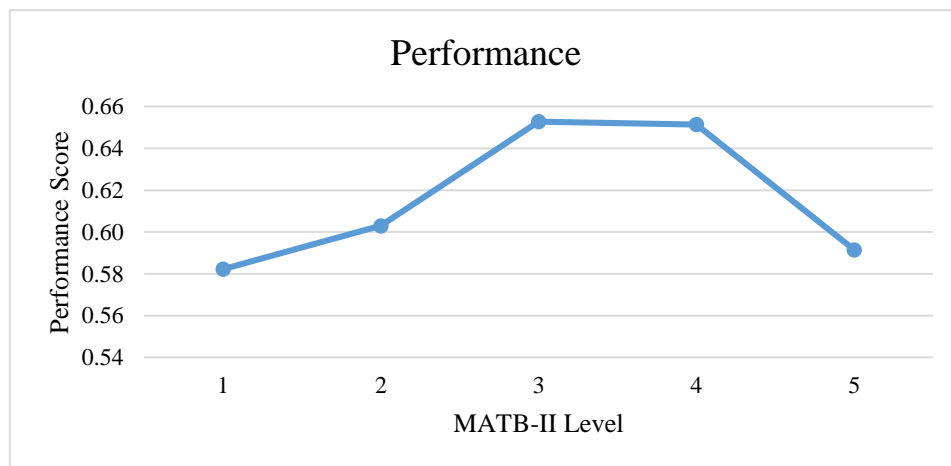
I_t = Number of correct responses with an incorrect answer

T = Time (in seconds) needed for the timeout to occur

E = Total number of events in the scenario.

The tracking task and resource management task were calculated using the percent difference from the target.

Performance resulted in a nearly inverted U-shaped graph, as seen in Graph 8: Performance. The results are provided by the percentage that the system is in the correct state. Level 1 scored an average of 58%, level 2 60%, level 3 and level 4 each scored an average 65%, while level 5 score 59% on average. There was no significance between the levels.



Graph 8: Performance

Relations Between Variables

Several methods exist to find correlations between variables, such as Chronback's alpha. To find the relationship between variables needed specific analysis, as all the variables that needed to be correlated are dependent in this study. Thus, a MANOVA with multiple dependent measures was performed. Wilk's Lambda demonstrated significant results between the three strongest physiological measures of workload (EEG (Alpha wave), SCR, and HRV). Wilk's Lambda: 0.772, $F(8, 182)=3.14$, $p=0.0024$.

Analysis

Of the different physiological measures, the most accurate one is HRV, where the plateau is visible. The problem with HRV is that it needs to be computed from the R-R (peak-to-peak) information provided by one of the devices. Some devices, such as BioHarness, can compute this automatically, but it is not always clear what time window the device uses for the computations, resulting in having a real-time output that might not correspond to a certain task.

Similarly, EEG results proved significant. Level 1 proved very similar to levels 2, and 3, but levels 4 and 5 are different from these first three levels. Levels 4 and 5 are near-identical to each other, indicating that the physiological measure is no longer changing.

Skin conductance response is another nearly-significant physiological measure to detect workload. Even though the graph shows a decrease for the fourth difficulty

level, statistical results show no significant difference between levels 3, 4, and 5, indicating a plateau, where the SCR stopped changing.

Subjective measures resulted in being very good indicators of mental workload. The plateau is very prominent in WPI, while in NASA-TLX the scores had stabilized for the difficulty levels 3, and 4, but continued increasing in level 5. This increase at the last level could also indicate increased effort by the users.

This increased user effort in levels 3, 4, and 5 could also explain why performance increased for levels 3 and 4, but ultimately dropped for level 5, as the user's resources were maxed out regardless of effort. Users could have detected the increased level difficulty, which made them more aware of what was happening in the level, increasing their arousal level, reflected in their increased performance.

Thus, based on these results, **hypothesis 1:** Physiological measures changes (decreasing or decreasing, depending on the measure) will correlate with level increase for the first few levels, while for the last levels marginal correlation with increased level will be found (when the redline is reached) can be proven true for HRV, SCR and EEG, where the correlation with the difficulty levels was marginal for the last 2 levels.

Hypothesis 2: Performance decrements do not necessarily correlate with the difficulty levels, as it was found not to be significant, as performance in levels 3 and 4 was higher than the first levels. A possible reason to explain this can be that the user was underloaded for the first two levels. Thus, further analysis is required to determine the exact cause.

CONTRIBUTIONS AND IMPACT

This project can contribute to the information on how to design better working environments where workers are under high workload. Being able to correctly identify the cognitive redline using real-time physiological measures can alert when a worker is nearing their maximum capacity, which can help to introduce safe workload levels. This can provide a clue as to when an additional worker or automation may be necessary in order to reduce the high workload. This not only helps with safety, but makes sure that performance can be kept at optimal levels as well.

A framework can be created by improving the methods described in this thesis and taking the knowledge out to real job settings. This framework can allow the ergonomist to potentially identify the cognitive redline before a new system is developed, thus making sure the system has safe workload levels before its implementation (Grier et al., 2008). This can provide assistance to multitasking jobs that incorporate high cognitive demands, which can hypothetically be kept at safe levels. The two main benefits include having the knowledge of how much mental workload is safe in certain high-profile systems, such as aircraft cockpits, surgery rooms, or factory assembly lines. This workload can also be analyzed through simulation software.

FUTURE RESEARCH

This research has low ecological validity, as it was performed in the lab with a software to induce mental workload. More research is needed in the area of the cognitive-red line of workload, as it still needs to be validated in an ecological way, such as studying it in the real workplace to validate safety and incident rates to high workload (Grier et al., 2008). Knowledge about the red-line could also help with automation, as workload can be determined and the task can be automated before the human operator reaches the red-line (Grier et al., 2008).

The study will continue using other physiological measures, such as electrooculogram (EOG). This added physiological measure can be compared with the others to see which one is more relevant as a real-time indicator of mental workload.

CONCLUSION

Being able to correctly identify the cognitive redline through real-time physiological measures can bring great benefits to any high-workload multitasking place. For example, safe levels of workload can be found and implemented, and new displays developed. In addition, when trying to introduce automation, knowing where the cognitive redline of workload lies can ensure that automation is introduced to alleviate workload and not add more tasks on the human operator.

Heart rate variability is the physiological measure that seems to be the most reliable, followed by EEG (Alpha wave) and skin conductance response. Knowing that these physiological measures can be used to monitor workload in real time can lead to a better assessment of workload in areas with a high task demand to increase user safety and performance.

Since this study took place in a laboratory setting, the low ecological validity but high experimental control served to prove the existence of the redline, as it was noted that some physiological measures reach a point where they no longer increase or decrease, regardless of added mental workload. More analysis is required to increase the ecological validity of the study, which can be done in follow-up studies.

REFERENCES

- AP. (2015). Indy 500 broadcast to introduce real-time telemetry to fans. Retrieved from <http://abcnews.go.com/Sports/wireStory/indy-500-broadcast-introduce-real-time-telemetry-fans-31242601>.
- Battiste, V., & Bortolussi, M. (1988). Transport pilot workload: A comparison of two subjective techniques. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 32(2), 150-154.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78 (Supplement 1), B231-B244.
- Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). United States: Springer Science and Business Media.
- Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, 42(3), 898-903.
- Brown, B. B. (1970). Recognition of aspects of consciousness through association with EEG alpha activity represented by a light signal. *Psychophysiology*, 6(4), 442-452.
- Chiappe, D., Conger, M., Liao, J., Caldwell, J. L., & Vu, K. P. (2013). Improving multi-tasking ability through action videogames. *Applied Ergonomics*, 44(2), 278-284.
- Cinaz, B., Arnrich, B., La Marca, R., & Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Personal Ubiquitous Computing*, 17(2), 229-239.

- Comstock, J. R., & Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*. Hampton, VA: National Aeronautics and Space Administration, Langley Research Center.
- Coyne, J. T., Baldwin, C., Cole, A., Sibley, C., & Roberts, D. M. (2009). Applying real time physiological measures of cognitive load to improve training. In D. Schmorow, I. Estabrooke, & M. Grootjen (Eds.), *Foundations of augmented cognition. Neuroergonomics and operational neuroscience* (Vol. 5638, pp. 469-478): Springer Berlin Heidelberg.
- Di Stasi, L. L., Antolí, A., Gea, M., & Cañas, J. J. (2011). A neuroergonomic approach to evaluating mental workload in hypermedia interactions. *International Journal of Industrial Ergonomics*, 41(3), 298-304.
- Dixon, S. R., & Wickens, C. D. (2006). Automation reliability in unmanned aerial vehicle control: A reliance-compliance model of automation dependence in high workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(3), 474-486.
- Fairclough, S. H., & Venables, L. (2006). Prediction of subjective states from psychophysiology: A multivariate approach. *Biological Psychology*, 71(1), 100-110.
- Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2), 171-184.
- Fréard, D., Jamet, E., Le Bohec, O., Poulain, G., & Botherel, V. (2007). Subjective measurement of workload related to a multimodal interaction task: NASA-TLX vs. Workload Profile. In J. Jacko (Ed.), *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments* (Vol. 4552, pp. 60-69): Springer Berlin Heidelberg.
- Galy, E., Cariou, M., & Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types?. *International Journal of Psychophysiology*, 83(3), 269-275.

- Grier, R., Wickens, C., Kaber, D., Strayer, D., Boehm-Davis, D., Trafton, J. G., & John, M. S. (2008). The red-line of workload: Theory, research, and design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(18), 1204-1208.
- Gurses, A. P., Carayon, P., & Wall, M. (2009). Impact of performance obstacles on intensive care nurses' workload, perceived quality and safety of care, and quality of working life. *Health Services Research*, 44(2p1), 422-443.
- Hancock, P. A., & Szalma, J. L. (2003). The future of neuroergonomics. *Theoretical Issues in Ergonomics Science*, 4(1-2), 238-249.
- Hancock, P. A., Williams, G., Manning, C. M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, 5(1), 63-86.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In A. H. Peter & M. Najmedin (Eds.), *Advances in Psychology* (Vol. Volume 52, pp. 139-183): North-Holland.
- Helton, W. S. (2004). Validation of a short stress state questionnaire. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(11), 1238-1242.
- van den Hombergh, P., Engels, Y., van den Hoogen, H., van Doremalen, J., van den Bosch, W., & Grol, R. (2005). Saying 'goodbye' to single-handed practices; what do patients and staff lose or gain? *Family Practice*, 22(1), 20 - 27.
- Jeong, E. J., & Biocca, F. A. (2012). Are there optimal levels of arousal to memory? Effects of arousal, centrality, and familiarity on brand memory in video games. *Computers in Human Behavior*, 28(2), 285-291.
- Karim, N., Hasan, J. A., & Ali, S. S. (2011). Heart rate variability - a review. *Journal of Basic and Applied Science*, 7(1), 71-77.
- Kim, J. Y., & Ji, Y. G. (2013). A comparison of subjective mental workload measures in driving contexts. *Journal of the Ergonomics Society of Korea*, 32(2), 167-177.

- Kohlmorgen, J., Dornhege, G., Braun, M. L., Blankertz, B., Müller, K.R., Curio, G. Kincses, W. E. (2007). Improving human performance in a real operating environment through real-time mental workload detection. *Toward Brain-Computed Interfacing*: MIT Press.
- Krehl, C., & Balfe, N. (2014). Cognitive workload analysis in rail signalling environments. *Cognition, Technology & Work*, 16(3), 359-371.
- Matthews, G. & Campbell, S. E. (1998). Task-induced stress and individual differences in coping. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(11), 821-825.
- Matthews, G., Emo, A. K., Funke, G., Zeidner, M., Roberts, R. D., Costa, Jr, P. T., Jr., & Schulze, R. (2006). Emotional intelligence, personality, and task-induced stress. *Journal of Experimental Psychology: Applied*, 12(2), 96-107.
- Matthews, G., & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition & Emotion*, 24(2), 357-373.
- Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, (2138), 6-12.
- Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers In Human Neuroscience*, 7.
- Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., & Yamamura, T. (2009). Physiological responses to workload change. A test/retest examination. *Applied ergonomics*, 40(6), 987-996.
- Modesti, P. A., Ferrari, A., Bazzini, C., Costanzo, G., Simonetti, I., Taddei, S., & Sirigatti, S. (2010). Psychological predictors of the antihypertensive effects of music-guided slow breathing. *Journal of Hypertension*, 28(5), 1097-1103.

- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32(1-2), 1-8.
- Phillips, J. B., & Boles, D. B. (2004). Multiple resources questionnaire and workload profile: application of competing models to subjective workload measurement. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 48(16), 1963-1967.
- Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *The International journal of aviation psychology*, 10(4), 393-410.
- Robert, G., & Hockey, J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological psychology*, 45(1), 73-93.
- Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration?. *Biological psychology*, 34(2), 259-287.
- Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology*, 53(1), 61-86.
- Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock Jr, J. R. (2011). The Multi-Attribute Task Battery II (MATB-II) software for human performance and workload research : A user's guide. *NASA Tech Memorandum*, 217164.
- Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., & Mehler, B. (2014). Classifying driver workload using physiological and driving performance data : Two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 4057-4066
- Splawn, J. M. (2013). *Applying Hyperspectral Imaging to Heart Rate Estimation for Adaptive Automation*. (Master of Science in Systems Engineering Thesis). Air

Force Institute of Technology, Wright-Patterson AFB OH Graduate School Of Engineering And Management. (No. AFIT-ENV-13-M-30)

TrexLLC (2014). Paced Breathing. (2.0). [Mobile application software]. Retrieved from <https://play.google.com/store/apps/details?id=com.apps.paced.breathing&hl=en>

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological psychology*, 42(3), 323-342.

Vidulich, M. A., & Tsang, P. S. (2012). Mental workload and situation awareness. In Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomics*, (3rd ed.), 243-273. John Wiley & Sons.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748-751.

Wang, Z., Hope, R. M., Wang, Z., Ji, Q., & Gray, W. D. (2012). Cross-subject workload classification with a hierarchical Bayes model. *Neuroimage*, 59(1), 64-69.

Waterhouse, J., & Campbell, I. (2014). Reflexes: principles and properties. *Anaesthesia & Intensive Care Medicine*, 15(4), 214-219.

Why locate the sensor at FP1? . (2014). Retrieved from: <http://support.neurosky.com/kb/science/why-locate-the-sensor-at-fp1>.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3), 449-455.

Wickens, C. D., Hollands, J. G., Parasuraman, R., & Banbury, S. (2012). *Engineering psychology and human performance* (4th ed.). Pearson.

Wilson, G. F., Caldwell, J. A., & Russell, C. A. (2007). Performance and psychophysiological measures of fatigue effects on aviation related tasks of

varying difficulty. *The International Journal Of Aviation Psychology*, 17(2), 219-247.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18(5), 459-482.

ZephyrCorporation (2008). Validity of BioHarness heart rate vs 3-lead ECG. Retrieved from <http://zephyranywhere.com/media/WhitePapers/WhitePaper-ZWP-003-Validity-of-BioHarness-Heart-Rate-vs-3-lead-ECG.pdf>